[0050] **Figure 4** is an illustration of an exemplary method of merging read pairs along a continuum. This merge can be performed by choosing one read pair, *e.g.*, read pair (3,61) from **Figure 3**. A direction is arbitrarily chosen for the first alignment, in this case, it is decided to represent the reads from left to right. Pair (3,61) is aligned along a continuum running from left to right. Next, a read pair can be chosen that includes one of the reads in the pair already aligned. Pair (3,50) can be chosen, *e.g.*, and read 50 aligned with read 3 in **Figure 4**. The merge can be built pair by pair as shown in **Figure 4**. It should be noted that read 14, shown in **Figure 4**, would not be aligned from the data shown in **Figure 3** because it does not share a common subsequence with the other reads shown. However, as the reads are merged with data not shown in **Figure 4**, other pairs can be aligned until a read is added that has a subsequence in common with read **14**, and read **14** would be merged. From the merged reads, the resolved sequence (SEQ ID No. 1) ATAGCCCTGCGCCTATCG, indicated below the continuum line in **Figure 4**, can be obtained.

[0051] Alignments optionally can use the associated position of the subsequences on the reads to confirm overlap. For example, consider the two sequences: GATCCCATGCGCA (SEQ ID No. 2) and ATAGCCCTATGAT (SEQ ID No. 3). These sequences share common subsequences GAT and CCC. We know from the associated position information that CCC begins at base 4 for the former and base 5 for the latter, and the GAT subsequence begins at base 1 for the former and base 11 for the latter. This position information indicates that there is no overlap between these two sequences since the position of the common subsequences does not allow for alignment. Consider now the first sequence above and the sequence CCCATGCGCATAT (SEQ ID No. 4). We know that the common subsequence CCC begins at base 4 for the former and base 1 for the latter, and the common subsequence CGC begins at base 9 for the former and base 6 for the latter. This position information indicates that there is overlap between these two sequences. Comparing the rest of the intervening subsequences confirms the overlapping region CCCATGCGCA (SEQ ID No. 5).

[0061] **Figure 7** is an illustration of an exemplary method of identifying a repeat region **R** and a set of unique regions **A, B, C, D**, in accordance with the present invention. By distilling the sequences of the merged reads, as shown in **Figure 7**, the portion of the merged reads that is not branched can be identified as repeat region **R** (SEQ ID No. 6), and the arms or branches are unique regions **A, B, C, D**. Only two unique regions are shown on each side of repeat region **R** for the sake of illustration. However, in practice a repeat region can be flanked by more unique regions depending on the number of times the repeat region appears in the genome.

[0064] **Figure 8** is an illustration of an exemplary method of linking pairs of unique regions using the linking information associated with the reads in the unique regions, and inserting the repeat region between each linked pair of unique regions with which the repeat region corresponds, in accordance with the present invention. The correct assembly of the unique regions can determined by searching or otherwise identifying reads in the unique regions with linking information to reads also in the linking regions. For example, **Figure 8** depicts a link between reads 111 and 62 ($\mathbf{r_{111}}$, $\mathbf{r_{62}}$), with a known orientation relative to each other and a known distance between reads, $\mathbf{d_{111,62}}$. It also shows reads 320 and 305 ($\mathbf{r_{320}}$, $\mathbf{r_{305}}$), with a known orientation relative to each other and a known distance between the reads $\mathbf{d_{320,305}}$. From this linking information, it can be determined that regions **A** and **D** flank one copy of repeat region **R** (SEQ ID No. 7), and regions **B** and **C** flank a second copy of repeat region **R** (SEQ ID No. 8). Once the linked pairs are identified, the method can optionally include inserting the repeat region between the linked pairs it corresponds to as shown in **Figure 8**. What is not known from the information depicted in **Figure 8** is how the linked pairs of unique regions are oriented relative to each other and at what distance.

In the Sequence Listing

Please amend the application to include the Sequence Listing submitted on even date herewith to BOX SEQUENCE. In accordance with 37 C.F.R. § 1.823(a), the pages of the attached Sequence Listing are numbered independently of the numbering of the remainder of the application.